

Javaslat a magyar igei WordNet kialakítására

Kuti Judit¹, Vajda Péter¹, Varasdi Károly¹

¹MTA Nyelvtudományi Intézet – Korpusznyelvészeti Osztály
1068 Budapest, Benczúr u. 33.
{kuti j, vajda, varasdi}@nytud.hu

Kivonat: Előadásunkban a magyar igei WordNet elkészítésének javasolt módszerét mutatjuk be. A Princeton WordNet automatikus fordításának kézi ellenőrzése során elvetettük a WordNet hierarchikus struktúrájának kritika nélküli átvételét, mivel az a jelentések elkülönítésében és a hierarchia kialakításában következetlenségeket tartalmaz, továbbá nem tükrözi a magyar igei lexikális viszonyait. Kiinduló jelentéshalmazunk meghatározásához ugyanakkor megtartjuk fogalmainak egy részét, és ezeket megkíséreljük a magyarban gyakori jelentésű igeikkel kibővíteni. Az igei közötti relációkat egy olyan általunk kialakított módszer alapján vesszük fel, amely nyelvészeti megalapozott szemantikai teszteken alapul, valamint megengedi, hogy egy szó több másikkal is alá-fölérendelt viszonyban legyen. A jelentések, relációk és az igei hierarchia kialakításához a Magyar Értelmező Kéziszótárt és több európai nyelvre elkészült wordnetet is felhasználunk.

1 Bevezetés

A magyar WordNet adatbázis a 2005 tavaszán indult *Magyar Ontológia Építése* projekt keretén belül készül, mely a Szegedi Tudományegyetem, a MorphoLogic Kft. és a Nyelvtudományi Intézet közös projektuma. A Nyelvtudományi Intézet a magyar EuroWordnet igei részének elkészítését vállalta magára. A EWN és a BalkaNet projektek (ld. [8], [2], [9], [10]) lezárulásával két módszer vált nemzetközileg elfogadottá WordNetek kialakítására. Az egyik az ún. *kiterjesztéses módszer* (*expand model*), melynek során a készülő WordNet mag-synset halmaza a Princeton WordNet 2.0-ból (a továbbiakban PWN) kerül ki. A kiválasztott PWN synseteknek célnyelvi synset-megfeleléseket keresnek, melyek megőroklk a PWN hierarchikus relációit. Ezt a nagyrészt automatizálható munkafázist manuális ellenőrzés követi, saját nyelvi erőforrások használatával. A másik módszer az ún. *összevonásos módszer* (*merge model*), melynek során saját erőforrásokból kiindulva határozzák meg a kiindulási jelentéshalmazt, és a közöttük fennálló relációkat is önállóan alakítják ki. A mostanra szabvánnyá vált PWN-tel való kompatibilitást az ezután következő munkafázis biztosítja, mely megfelelteti a készülő WordNet synsetjeit a PWN aktuális verziójának synsetjeivel. A HuWN kialakítását eredetileg a melléknévi és főnévi részhez hasonlóan, túlnyomó részben a kiterjesztéses módszerrel terveztük. Ez a BalkaNet projektben használt nyelvek közötti közvetítő réteg (Balkanet Concept Set (BCS)),

mely 8516 PWN synsetnek megfeleltetett jelentést tartalmaz, ebből 2318 igei synset) automatikus módszerekkel történő magyar nyelvre képezését, majd manuális ellenőrzését és kiterjesztését jelentette volna. A manuális ellenőrzés kezdeti fázisában felmerülő problémák ill. kérdések, valamint az az igény, hogy saját nyelvi erőforrásaink teljesebb kihasználása érvényesüljön, olyan módszertani változtatásokat, kiegészítéseket tett szükségessé, melyek jobban figyelembe veszik mind az igei szófaj, mind a magyar nyelv lexikalizációs sajátosságait.

A következő pontban megvizsgáljuk a Princeton WordNet azon tulajdonságait, melyek fogalmainak és hierarchiájának teljes átvétele ellen szólnak, és megindokoljuk a kiterjesztéses modell elvetését. A 3-ik pontban megemlíjtük az igei sajátosságaiból adódó szempontokat, melyeket a HuWN kialakításakor figyelembe veszünk. Bemutatjuk a javasolt munkamódszert, kitérve a magyar WordNetbe felveendő jelentések kiválasztására és a köztük lévő relációk kialakítására. Az utolsó pontban a további munkafázisokról ejtünk néhány szót.

2 A Princeton WordNet átvételének korlátai

A felmerülő problémák részben a kiindulási alapot képező PWN hiányosságaiából (ld. alább, ill. részletesebben [3]), részben pedig a tervezett kiterjesztéses módszer korlátaiból adódtak.

2.1 A jelentések elkülönítése

A kiterjesztéses módszer alkalmazásával a PWN belső hierarchiáját a magyar hierarchia is megörökölné. A PWN igei részének azonban számos olyan jellemzője van, amelyeknek az átvételét az igei HuWN készítésekor el szeretnénk kerülni, ill. amelyet módosítani szeretnénk. Alapvető kérdésként merül fel, hogy a PWN felépítése során mely elv határozta meg a jelentések megkülönböztetését. Ha abból indulunk ki, hogy a PWN *nyelvi ontológia* (ld. [10]), azaz egy bizonyos nyelv *lexikalizálódott* jelentés-megkülönböztetéseit hivatott tükrözni, nem egyértelmű, mi alapján kerültek egy ill. több synsetbe a következő jelentések:

*{shed:4, molt:1, exuviate:1, slough:1, moult:1}*¹⁶ (def.: Némely állat bizonyos időszakban leveti kültakarója bizonyos rétegét, pl. szőrét, tollát, agancsát.)

{feed:2, give:24} (def.: enni ad valakinek) ill. *{give:19}* (def.: gyógyszerrel ad valakinek)

{clean:1, make clean:1} (def.: valamit a rajta vagy benne levő szennytől egészen megtisztít) ill. *{wash:3, launder:1}* (def.: Vízzel tisztít, általában ruhaneműt.)

Míg a *{shed:4}* synset szinonimái közül a *shed* minden típusú levetett állati külsőre vonatkozhat, a *slough* viszont nem vonatkozhat madártollra, csak kételtű vagy hulló bőrére ill. agancsra, addig a *clean:1* és a *{wash:3, launder:1}* között (bár a definíciók

¹⁶ Kapcsos zárójelbe kerül egy synset, amennyiben több elemével utalunk rá.

ezt nem tükrözik egyértelműen) pont a tárgyi vonzat milyensége adja meg a különbséget: a *{clean:1}* bármire vonatkozhat, a *launder:1* azonban csak ruhaneműre. Ennek ellenére ugyanabba a synsetbe tartozik a *shed:4* és *slough:1*, ellenben két külön synsetbe a *clean:1* és a *launder:1*. A *give:24* és *give:19* között fellelhető egyetlen különbség szintén csupán abban áll, hogy a tárgyi vonzat ennivaló vagy gyógyszer-e, de annak ellenére, hogy a köztük levő különbség nem lexikalizálódott, két külön synsetbe kerültek. Ezek a példák azt mutatják, hogy a PWN nem konzekvens a lexikalizálódott vonzatok kezelésének tekintetében. Amennyiben nyelvi ontológiának tekintjük a PWN-t, szintén megkérdőjelezhetővé válik a következő synseteknek, mint lexikalizálódott kifejezéseknek a felvétele a hierarchiába: *create from raw material:1*, *create from raw stuff:1*, *change magnitude:1* ill. *change integrity:1*. Amennyiben azonban a PWN nem „csupán” nyelvi ontológiának tekintendő, hanem legalább részben következtetés alapú ontológiának (ld. [10]), érthetővé válik ezeknek a csomópontoknak a bevezetése, de szembeötlő mesterséges, azaz nem lexikalizált csomópontokként való megjelölésük hiánya.

Szintén inkonzekvensnek hat némely metaforizációra visszavezethető jelentésmegkülönböztetés.

(i)

{take away:1, carry away:1} def.: valamely helyről, környezetből, mentális vagy érzelmi állapotból kimozdít – usage: The car carried us off to the meeting; I got carried away when I saw the dead man and I started to cry.

{sweep:1, brush:4} def.: át- vagy keresztülsöpör – usage.: Her long skirt brushed the floor.; A gasp swept cross the audience.

(ii)

take off:3 def.: eltávolodik a földtől – usage.: The plane took off two hours late.

take off:7 def.: elindul, mozgásba jön *átvitt értelemben* – usage.: the *project* took a long time to get off the ground.

Mivel az (i) alatt szereplő szinonimák használatára vonatkozóan metaforikus és nem metaforikus példák is szerepelnek, indokolatlannak tűnik a (ii) alatti *take off* két külön jelentésének megkülönböztetése figuratív használat ürügyén.

2.2 A fogalmi hierarchia átvételének hátrányai

Az előző pontban említett módszertani következtetlenségeken kívül számos olyan eset van, ahol a hierarchia kialakítása mögött álló elv megvalósítása hibásnak tűnik. Ilyen például az ágens - páciens tematikus szerepek alternációjának inkonzekvens kezelése. Például több ágens alanyú ige (*correct*, *falsify*, *undo*, *modify*) a páciens alanyú, és nem az ágens alanyú *change* hiponímájaként (azaz a *változtat* helyett a *változik* csomópont alatt) szerepel.

Ezektől a hibáktól eltekintve is bizonyos negatívumokkal járna az angol fogalmi háló teljes átvétele. Egyrészt elkerülhetetlen, hogy kevésbé fogja tükrözni a magyar lexikalizációs viszonyokat az elkészült HuWN, mintha saját erőforrásokra támaszkodva vennénk fel a relációkat. Másrészt köztudomású, hogy a PWN hierarchia

csomópontjai olyan sűrű fogalmi hálót alkotnak, hogy magyar nyelvre történő leképezésük nemcsak sok esetben manuálisan is erőltetettnek bizonyul, de az alábbi nehézségeket is magában hordozza.

Ha a tervezett módszert követve, először a PWN összes, a BalkaNet Base Concept Set-jét alkotó synsetjét megfeleltetjük magyar synseteknek, majd azokat feleltetjük meg ÉKSz-beli jelentéseknek, óhatatlanul lesznek olyan ÉKSz jelentések, melyek kimaradnak egy adott ige jelentései közül. Ezeket ugyan egy későbbi, a BalkaNet Concept Set-et bővítő munkafázis során bevehetjük a HuWN-be, de nem lesz rá lehetőségünk, hogy esetleg már synsetként felvett jelentésekkel összevonjuk őket. A tapasztalat pedig azt mutatja, hogy az ÉKSz rengeteg olyan kollokációt és idiomatikus kifejezést tartalmaz, amelyeket a HuWN céljainak megfelelően össze lehetne vonni kevesebb synsetbe is.

A PWN fogalmi sűrűségének magyarra való leképezése azért is lenne problematikus, mert a készülő HuWN egyik legfontosabb felhasználása egy információkinyerő rendszer hatásfokának javítása lesz. Amennyiben azonban a HuWN igei része átveszi a PWN fogalmi hálóját, túlságosan sűrű lesz ahhoz, hogy hatékony segítséget nyújtson nyelvfeldolgozási alkalmazásokban. A PWN a *go* lemmának például 28, a *cut* lemmának 21, a *see* lemmának pedig 19 igei jelentését különbözteti meg. Többnyelvű információkinyerés esetében exponenciálisan nő a keresés eredményekor a zaj, ha az eredeti keresett kifejezéshez hozzávesszük a PWN synsetjeiből való szinonimákat. Ha csökkenteni tudjuk a többértelműséget, pontosabb eredményeket fog adni egy-egy keresés (ld. [7]).

Végül meg kell említeni, hogy a PWN készítése során nem alkalmazták a többszörös öröklődés lehetőségét a hierarchiában. Minden fogalomnak fix helye van, amely meghatározza jelentését. Bár ez a módszertani döntés érthető, hiszen a PWN készítése a főnévi szókincs építésével kezdődött, az igei HuWN készítésénél úgy gondoljuk, hasznos lenne bizonyos esetekben a többszörös öröklődés bevezetése.

3 Javasolt módszertani változások

3.1 A kiinduló jelentéshalmaz meghatározása

Az igeik, mint eseményszerűségeket leíró egységek, általában nyelvspecifikusabb módon tükröznek lexikalizációs mintákat, mint a főnevek, hiszen ugyanannak az eseménynek több szempontját is megnevezhetik. Pl. az angol *clear:24* (def.: megtisztítani a torkot valamitől, rekedtes hangadás kíséretében) synset a *remove:1* (def.: valami konkrétat eltávolít valahonnan, emeléssel, nyomással stb., vagy valami absztrakt dolgot eltávolít) hiponímája az angolban, míg az ennek a magyarban megfelelő (meg)köszörrüli (a torkát) inkább valamilyen hangadást kifejező ige hiponímája lehetne. Ebből adódóan jobban szeretnénk függetleníteni a HuWN igei részét az angol nyelv által meghatározott fogalmi hálótól, mint a főnévi és melléknévi részt.¹⁷ Ennek

¹⁷ A főnévi és melléknévi HuWN kialakítása a következő módszertan szerint történik: a PWN fogalmi csomópontjait automatikus módszerekkel megfeleltetjük magyar synset-kezdeményeknek, és átvesszük a közöttük fennálló relációkat. Az automatikus fordítás munkafázisát alapos manuális ellenőrzés követi.

érdekében három fontos szempontot tartanánk szem előtt az igei HuWN kialakításakor:¹⁸

- (i) Csak PWN fogalmi csomópontokat képezünk le magyarra, a köztük fennálló relációkat nem.
- (ii) Korlátozzuk az angolból automatikus leképezéssel átvett igei jelentések számát a taxonómiai fontos jelentésekre.
- (iii) Az automatikus leképezéssel nyert kiinduló fogalmak körét kiegészítjük saját erőforrásokból származó jelentésekkel.

Taxonómiai fontos jelentésűnek vesszük azokat a PWN-beli jelentéseket, melyek magas hierarchiai pozícióval rendelkeznek. Ezt a feltételt teljesítik a EWN igei base concept-jeiből nyert BCS synsetek (254 igei synset) és a BCS igei felső csomópontjai. Ezeket a synseteket egészítjük ki olyan igei jelentésekkel, melyekről feltételezhető, hogy gyakoriak, ill. hogy releváns szerepet játszanak lexikális ismeretek tárolásakor. Mivel jelentésgyakorisági adatok nem állnak a rendelkezésünkre, az előbbi szempont figyelembevétele érdekében igei lemma gyakorisági lista készült a Magyar Nemzeti Szövegtárból, míg az utóbbi szempont betartásának érdekében két igei lemma gyakorisági lista készült a Magyar Értelmező Kéziszótár elektronikus verziójából: egy igei genus proximum gyakorisági lista az igei definíciókból, és egy igei lemma gyakorisági lista az összes ÉKSZ definíciókból.¹⁹ Az MNSZ-ből és az ÉKSZ összes igei definíciójából kapott lemma gyakorisági lista első 50 elemének metszetét kiegészítettük a genus proximumok listájának azon elemeivel, melyek ebben a metszetben még nem szerepeltek. Az így kapott lista 28 elemet tartalmaz, melyeknek összesen 471 ÉKSZ-beli jelentés felel meg. A Nyelvtudományi Intézet birtokában levő igei vonzatkeret adatbázis rekordjai közül az említett 28 lemmának 371 vonzatkeret felel meg. Terveink szerint ennek a 471 ÉKSZ-beli jelentésnek, és 371 igei vonzatkeretnek azon elemeit fogjuk első lépésben hierarchiába rendezni, amelyek angol megfelelői a BCS részét képezik. Az így kapott magyar fogalmi hálóbba fogjuk ezek után második lépésben beleilleszteni a még nem lefedett BCS igei jelentéseit. Ilyen módon, bár a HuWN igei része maximálisan kompatibilis marad mind a Princeton WordNettel, mind a BalkaNettel (hiszen a BCS igei részét megfeleltetjük PWN synseteknek), mégis jobban fogja tükrözni a fogalmi háló a magyar nyelv lexikalizációs sajátosságait, mintha kizárólag angol irányból közelítettük volna meg a mag-wordnet készítését.

Lehetővé válik, hogy indokolt esetben (ld. 2.2.) több ÉKSZ-beli jelentést összevonva alakítsunk ki magyar synseteket, illetve, hogy bevezessük a többszörös öröklődést a hierarchiában. Ahogyan utaltunk már rá, a PWN nem alkalmaz sem többszörös öröklődést, sem mesterségesnek megjelölt csomópontokat. Ennek ellenére az igei szófaj sajátosságai miatt tervezzük bevezetni ezeket a lehetőséget is. Bizonyos morfémák a magyarban (igekötők ill. képzők) kifejezhetnek akcióminőséget ill. az igével kifejezett állapot változását (ld. [6]). Ezeket az eseteket célszerű lenne a több-

¹⁸ A következőkben leírtakat a EWN résztvevőinek módszertani döntéseire alapozva alakítottuk ki (ld. [1]).

¹⁹ A gyakorisági listák elkészítéséért Miháلتz Mártont, Nagy Viktort és Oravecz Csabát illeti köszönet.

szörös öröklődés és a mesterséges csomópontok lehetőségét kihasználva elhelyezni a kialakítandó hierarchiában.²⁰ Így például az általában *el-*, *fel-* és *meg-* igekötős inchoatív akcióminőséget kifejező igék az alapige jelentése szerint és akcióminőségük szerint is kapnának hipernímát. A *felkacag* ige ugyanúgy lenne a mesterséges *kezd* akcióminőséget jelző csomópont hiponímája, mint a *kifejez*, *kimutat*-é.

3.2 A felveendő relációk típusai

A felveendő relációkat a PWN relációtípusaiból kiindulva határoztuk meg (ld. [4]).

Szemantikai és időbeli relációk

Az alapvető szemantikai reláció igejelentések között az *implikáció*: V_1 akkor implikálja V_2 -t, ha mindahányszor, amikor $X V_1$ igaz, $X V_2$ -nek is igaznak kell lennie, függetlenül X megválasztásától. Például: a *töpreng* implikálja a *gondolkodik*-ot, mert ha valaki töpreng, akkor az a valaki gondolkodik is. Hasonlóképpen, a *vásárol* implikálja a *fizet*-et, hiszen ha valaki vásárol, akkor fizet is. Ezt a viszonyt általánosan a következő sémával tesztelhetjük:

V_1 implikálja V_2 -t, ha abból, hogy „ $X V_1$ -zett”, következik, hogy „ $X V_2$ -zött”.

Ha valaki töprengett, akkor gondolkodott (is), illetve, ha valaki vásárolt, akkor fizetett (is). Implikációs viszony fennállhat időben diszjunkt események között is: ha valaki elveszített valamit, akkor – szükségképpen – előzőleg birtokolnia (is) kellett azt a valamit; ez a birtokviszony azonban az elvesztés pillanatában megszűnik, így a két esemény időben diszjunkt.

A PWN felfogása szerint az igék eseményeket írnak le. Pontosítva: minden V_i igehez (vonzatkeretének kitöltése után) hozzárendelhető egy e_i esemény, amelyet a szóbanforgó kijelentés leír. Mivel minden esemény időben zajlik, tetszőleges e eseményhez hozzárendelhető az esemény *futási ideje*, $\tau(e)$, ami általában egy időintervallum. Két esemény között az alábbi temporális viszonyok relevánsak a WN szempontjából:

- kotemporalitás: e_1 kotemporalis e_2 -vel, ha $\tau(e_1) = \tau(e_2)$;
- időbeli tartalmazás: e_1 időben tartalmazza e_2 -t ha $\tau(e_1) \subseteq \tau(e_2)$ (azaz e_1 minden pontja egyben e_2 -nek is pontja);
- szigorú időbeli tartalmazás: e_1 időben szigorúan tartalmazza (e_2) -t ha $\tau(e_1) \subsetneq \tau(e_2)$, de $\tau(e_1) \neq \tau(e_2)$ (azaz e_1 -nek van olyan pontja, amely e_2 -nek nem pontja);
- időbeli megelőzés: e_1 időben megelőzi e_2 -t, ha $\tau(e_1) < \tau(e_2)$ (azaz: e_1 minden pontja korábbi, mint e_2 bármely pontja).

²⁰ A mesterséges csomópontok rendszerét a GermaNet-re támaszkodva (ld. [5]) alakítjuk ki.

Szinonímia

V_1 **szinonim** V_2 -vel, ha

1. V_1 és V_2 kölcsönösen implikálják egymást és
2. szükségképpen kotemporálisak.

Részesemény

V_1 (valódi) **részeseménye** V_2 -nek, ha

1. V_1 implikálja V_2 -t és
2. e_1 szükségszerűen (szigorú) időbeli része e_2 -nek.

Például: *horkol* - *alszik*, *álmodik* - *alszik*, *összead* - *számol*

Troponímia

A troponímia a főnévi hiponímia igék közötti megfelelője. Lényegében azt a viszonyt jelöli, amikor egy esemény egyfajta *módja* egy másik esemény kivitelezésének.

V_1 egy **troponímája** V_2 -nek, ha

1. e_1 szükségszerűen időbeli része e_2 -nek és
2. „X V_2 -zött és közben V_1 -zett” mondat *nem* jólformált.

Például: *gitározik* - *zenél*, *vánszorog* - *jár*, *prédikál* - *beszél*

Temporális prekondíció

V_1 **temporális prekondíciója** V_2 -nek ha

1. e_2 implikálja e_1 -t és
2. e_1 időben szükségképpen megelőzi e_2 -t.

Például: *birtokol* - *elveszt*, *alszik* - *felébred*, *elalszik* - *alszik*

Okozás

Az okság alapvetően nem nyelvi kategória. A nyelvészetben azonban elfogadottá vált egy, a filozófiaival többé-kevésbé összeegyeztethető értelmezés, amit a PWN is felhasznál. E (nyelvészeti értelemben vett) oksági viszonynak a magyarban (is) van grammatikalizált megjelenése, a kauzatív alternáció. Ez azonban formailag meglehetősen szabályos és kiszámítható, így most nem foglalkozunk vele külön. A WN-ben releváns okságfogalom a példáik alapján részben a temporális prekondíció „tükrképe”, de vannak még további feltételek is:

V_1 oka V_2 -nek ha

1. e_1 implikálja e_2 -t és
2. e_2 nem előzheti meg e_1 -et és
3. e_2 függ e_1 -től abban a kontrafaktuális értelemben, hogy ha e_1 nem lenne (nem történné meg/abbamaradna), akkor e_2 sem lenne (nem történné meg / abbamaradna)

Így például, ha valaki *forogat* valamit, akkor ez azt okozza, hogy az a valami *forog*. A forgatás implikálja a tárgy forgását, a két esemény kotemporális, és fennáll a kontrafaktuális függés is (ha ugyanis a dolog anélkül is forogna, hogy az ágens ezért bármit tenne, nem mondanánk, hogy forgatja az illető dolgot).

A kontrafaktuális feltétel meglehetősen lényegi, mert ez akadályozza meg, hogy a természetes nyelvi okság tranzitív legyen.

Abból, hogy

- (1) Mari belakatolta az ajtót, mert zajt hallott.
és
(2) A zajt Mari vőlegényének közeledése okozta.

nem következtetünk arra, hogy

- (3) Mari belakatolta az ajtót, mert közeledett a vőlegénye.

Sajnos a PWN példái azt mutatják, hogy okságon a szerzők elég gyakran pusztán az első két feltétel teljesülését értik. Így pl. az *ad - birtokol* is oksági viszonyban áll, holott valaki anélkül is egy dolog birtokába juthat, hogy azt adták volna neki (és akkor már a *megvesz - birtokol* párt is oksági viszonyban állónak kell tekinteni, és ott is ugyanaz a helyzet).

Az igék közötti okozási viszony legfeltűnőbb sajátossága, hogy az okot leíró ige *alanya* leggyakrabban a másik ige *tárgyának* változását idézi elő (de: *megvesz - birtokol*). Ha pl. X lelövi Y-t, akkor Y az, aki átmegy az ige jelentésében implicite kódolt állapotváltozáson (ő hal meg). Általában: ha V_1 és V_2 oksági viszonyban áll, akkor

1. „X V_1 -te Y-t” implikálja, hogy „Y V_2 -zött” és
2. X ágens, Y páciens thematikus szerepű.

Például: *fékez - lassul, lelő - meghal, gyújt - gyűlik*

Antonímia

Az antoníma reláció nem synsetek, hanem különböző synsetek elemei között fennálló lexikális reláció a PWN-ben. Mivel tapasztalataink szerint az antonímia relációk meglehetősen hűséggel átvehetők az angol eredetiből fordítás útján, így ezeket megőrizzük a HuWN-ben.

4 További feladatok

A további munkafázisok során, a HuWN információkinyerő rendszerben történő alkalmazására való tekintettel a következő irányelveket szándékozunk követni: Annak érdekében, hogy az információkinyerő rendszer eseményleíró kereteken alapuló felismerőképessége javuljon, minden további felvett jelentéshez hozzárendeljük a megfelelő igei vonzatkeretet, ami segít az események szereplőinek meghatározásában. Ezenkívül megpróbáljuk integrálni a HuWN igei részébe az EWN-ben kidolgozott szemantikai attribútumok rendszerét, az ún. Top Ontology-t, mely a hierarchiát az igékhez rendelhető szemantikai jegyekkel egészíti ki.

Bibliográfia

1. Alonge, A.: Definition of the links and subsets for verbs. Deliverable D006. Technical Report WP4.1, EuroWordNet, LE2-4003, Amsterdam (1996)
2. Christodoulakis, D.: Balkanet. Deliverable D2.1. Technical Report WP 2. BalkaNet IST-2000-29388 (2000)
3. Cristea, D.: Mapping Princeton WordNet Synsets onto Romanian Wordnet Synsets. Romanian Journal of Information Science and Technology 7 (2004) 124–145
4. Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press (1998)
5. Hamp, B., Feldweg, H.: GermaNet — A Lexical-Semantic Net for German. In Vossen, P., ed.: Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications. Association for Computational Linguistics, New Brunswick, New Jersey (1997) 9–15
6. Kiefer, F.: Jelentélmélet. Corvina, Budapest (2000)
7. Kunze, C.: Semantics of Verbs within GermaNet and EuroWordNet (1999)
8. Tufis, D.: Balkanet: Aims, methods, results and perspectives. Romanian Journal of Information Science and Technology 7 (2004) 1–35
9. Vossen, P.: The EuroWordNet Base Concepts and Top Ontology, Deliverable D017, D034, D036. Technical Report EuroWordNet (LE 4003), University of Amsterdam, Amsterdam (1998)
10. Vossen, P.: EuroWordNet General Document. Technical Report EuroWordNet (LE2-4003, LE4-8328) (2005)